DOCUMENT RESUME

ED 333 029 TM 016 528

AUTHOR Yamamoto, Kentaro; Muraki, Eiji
TITLE Non-Linear Transformation of IRT Scale To Account for the Effect of Non-Normal Ability Distribution of the Item Parameter Estimation.
PUB DATE Apr 91
NOTE 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Ability; *Estimation (Mathematics); Graphs; *Item Response Theory; Mathematical Models; Maximum Likelihood Statistics; Simulation; *Statistical Distributions
IDENTIFIERS Ability Estimates; Item Parameters; National Assessment of Educational Progress; *Nonnormal Distributions

ABSTRACT

The extent to which properties of the ability scale and the form of the latent trait distribution influence the estimated item parameters of item response theory (IRT) was investigated using real and simulated data. Simulated data included 5,000 ability values randomly drawn from the standard normal distribution. Real data included the results for a random sample of 2,000 examinees on the age 13 mathematics trend assessments of the 1986 and 1988 National Assessments of Educational Progress. Results suggest that the prior distribution in marginal maximum likelihood estimation should be as close to reality as possible. Because of the lack of knowledge of the real ability distribution, it may be necessary to simultaneously estimate the multinomial ability distribution with the item parameters. When used with real data, this method is often unstable and sensitive to the initial values. A two-step procedure is recommended. First, converged model parameters are found using the normal prior distribution. Second, with these converged values as initial values, the multinomial ability distribution is estimated. Implications for IRT are discussed. Three tables and 10 graphs supplement the discussion. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED333029

NON-LINEAR TRANSFORMATION OF IRT SCALE TO ACCOUNT FOR THE EFFECT OF NON-NORMAL ABILITY DISTRIBUTION ON THE ITEM PARAMETER ESTIMATION[1]

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Kentaro Yamamoto and Eiji Muraki
Educational Testing Service

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
KENTARO YAMAMOTO
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ABSTRACT

The effects of a family of non-normal ability distribution on the IRT parameters estimated using the normal prior distribution are non-linear and can be systematic, this non-linearlity makes the application of the conventional method of linear transformation not completely satisfactory. A non-linear transformation of IRT scale to maximize fit and consequential item parameters using simulated data as well as real data is being described.

Introduction

In item response theory, the scaling and location of the latent trait distribution and resultant item parameter estimates have a reciprocal relation. If the scale of the latent trait distribution is multiplied by the factor k, the values of slope parameters will be the true value divided by k. By the same token, when the center of the latent trait distribution is shifted by the factor k, the values of the threshold parameters will be moved to the opposite direction to compensate for this shift. Many adjustment methods, including the item parameter equating method, utilize this linear relationship between a set of parameter estimates and the latent trait distribution. In other words, multiple sets of parameter estimates can be directly compared only when those estimates of parameter estimates assume that the latent trait distribution is on a interval scale. However, few researchers have discussed the effect of non-normal latent trait distributions on item parameter estimates. More generally, we investigated the extent to which properties of the ability scale and the form of the latent trait distribution influence the estimated item parameters of Item Response Theory.

Braun (1988) and others cautioned against relying on the implicit assumption of IRT ability scale being an interval scale. Mislevy (1987) also stated that only ordinal scale properties can be claimed for a scale derived from an application of IRT, with the exception of the Rasch

[1] Presented at the 1991 AERA meeting at Chicago.

TM016528

BEST COPY AVAILABLE

model. However, Lord (1980) stated that when model assumptions are satisfied, true item parameters do not change even when parameters are estimated from examinees with different true abilities sampled from the same population. Likewise, true abilities do not change, even when estimated from different sets of items.

Birnbaum (1968) discussed transformation of the ability scale, resulting in the existence of item characteristic curves for a given ability scaling. This means the item parameters can not be determined uniquely without considering the true latent trait distribution. Recently, Ramsay and Abrahamowicz (1989), and Ramsay (1989) developed an item response model using spline functions to capture the characteristics of an item more accurately than the conventional parametric item response models. Because of the lack of identifiability of the latent trait distribution and item parameters, they focused on estimating more accurate item response function by spline. There has not been a conclusive study of the lack of identifiability.

Perspective and Theoretical Framework

Item Response Theory (IRT) methodology produces an indeterminacy of the scale which stems from the parameterization of the model itself. The conventional method to "anchor" the scale is to standardize the ability distribution, i.e., set the mean and standard deviation of the ability distribution be 0.0 and 1.0, respectively. The scale of the item parameters follows once the ability distribution is standardized. The default option in BILOG follows this convention. Hence, the mean and standard deviation of the estimated ability distribution can be treated as the scaling constants for the item parameters. This linear transformation works well as long as the ability distributions have the same shape.

While the joint maximum likelihood method is free from the assumption on the shape of the ability distribution, the application of the marginal maximum likelihood method in BILOG requires the prior population distribution. In the past, several researchers reported Monte-Carlo studies on the relationship between the shape of true ability distribution and estimated item parameters. The results indicate that the skewness of the ability distribution interacts with the estimated item parameters. Yet, quantification of the bias and the discussion of how to minimize the effects have not been reported.

Methodology and Data Source

The initial idea concerning the interaction of ability distribution and estimated item

parameters is as follows. There are true and unknown ability distributions and also true/unknown item parameters on a common scale, and we only have an observable response data set. Our task is to estimate the ability distribution and the item parameters from the data set. At this point we have two choices, either 1) we estimate the multinomial distribution of ability or 2) use a prior distribution. The present study used the second option for the most part and the first option was used to evaluate the methods.

The standard normal distribution is most often used as a prior distribution. Obviously the choice of the prior distribution should not be treated lightly. It is best to use the true distribution if it was known. We introduced a scale transformation function between the true ability distribution and the prior ability distribution. Our intention is to propagate this transformation function to the item parameters. The prior distribution has a stronger influence on item parameter estimates, especially when the data provide less information about the shape of ability distribution. This is most noticeable when very few items are being calibrated. However, a larger number of items would result in the conditional likelihood being concentrated in a very small region for a response pattern. The smaller posterior variance yields a sum of the posterior distribution over the entire sample which is very different from the prior distribution.

Study 1:

The log-normal distribution was selected to represent the ability distribution. Since the log-normal distribution is the logistic transformation of the scale to the normal distribution, we can create the number of log-normally distributed data sets from one set of data which was generated from the standard normal distribution using the inverse transformation; this enables us to focus only on the transformation function of the scale.

The following log-normal distribution function is used

$$f(\theta,\gamma,\mu,\sigma^2)=\frac{1}{(\theta-\gamma)\sigma\sqrt{(2\pi)}}\exp\left(\frac{-[\ln(\theta-\gamma)-\mu]^2}{2\sigma^2}\right)\ \sigma^2>0,\ \gamma<\theta<\infty,$$

$$=0, \textit{otherwise}$$

The corresponding inverse transformation function for the normal scale to the log-normal scale is

$$\theta^t = \frac{\ln(\theta - \gamma) - \mu}{\sigma} \sim N(0,1)$$

The item characteristics function on the θ^t scale is

$$P(x=1|\theta^t, b, \sigma, \gamma, \mu) = [1 + \exp(-(\exp(\theta^t \sigma + \mu) + \gamma - b))]^{-1}$$

It is evident in the above equation that the item characteristic curve on the transformed scale does not have an ordinary IRT function, and the parameter estimation under adherence to the logistic IRT function yields non-linearly transformed parameters. Five levels of skewness were considered, each having the mean and standard deviation of 0.0 and 1.0, respectively. The values which specify those distributions are as follow.

Parameter Values of Log-Normal distribution used for the study

Set	σ	μ	γ
1			normal distribution
2	0.20	1.58	-4.95
3	0.40	0.80	-2.40
4	0.60	0.24	-1.52
5	0.80	-0.27	-1.06
6	1.00	-0.77	-0.76

Characteristics of Simulated Data

Five thousand ability values were randomly drawn from the standard normal distribution. Five levels of log-normal transformations were applied to the 5000 normally distributed abilities, hence 6 sets of data including the original data set without any transformation were obtained. Based on these ability values, the item parameter responses were generated. Each ability data set differ only by the transformation. That is to say that the random error of generated ability values of the parent distribution is equivalent across all data sets. These six ability distributions are plotted in figure 1. The one parameter logistic IRT model was selected for the clarity of the

scale/slope relationships and the ease of evaluation of estimation. The location parameters for the total of 17 items were equally spaced between -2.0 and 2.0.

The item parameters were estimated by BILOG (Mislevy and Bock, 1983) using the two parameter logistic IRT. The number of item parameters were increased from 1 to 2, because the transformation of the scale directly affect the value of slope. Two different options in regard to the type of constraints for the shape of the ability distribution were taken for the BILOG runs. One is to keep the standard normal quadrature weights until convergence for presumed normal ability distribution (NAD), and the other is to update the weights after each iterative step, thereby estimating a multinomial ability distribution (MNAD).

We performed one replication using exactly the same procedure. The obtained results were so similar, we decided to report results from on edata set. However, the statement applies to both replications.

Results (study 1)

The standardized (based on the mean and standard deviation of the posterior ability distribution) item parameters using both methods are plotted in Figures 2 through 5. Under both ability distribution constraints, the bias on the estimated item parameters are related to the skewness of the ability distribution. This was true for the both sets of item parameters. One notable observation is that the amount of bias is not uniform over the range of ability. The bias is greater where the transformation is greater. This type of systematic bias cannot be compensated by any type of linear transformation. The bias is less severe when a multinomial distribution is estimated. The standardized (mean = 0.0 amd s.d. = 1.0) posterior abiltiy distributions under two types of constraints are plotted in Figures 6 and 7. The posterior distributions of the MNAD method was much closer to the original distribution compared to the results from the NAD method. Regardless of the skewness of the original distribution, the posterior distribution resulting from the NAD method was similar to that from the standard normal prior distribution.

Often the method of evaluating the fit of the IRT model to a data set is the chi-square test at item level and -2log-likelihood value at the model level. The item level chi-square tests were non informative, i.e., all estimated item parameters fit well. However, the values of -2log-likelihood in Table 1 reveals some interesting result to evaluate the fit of the model. At first glance, the MNAD estimation was always found to be better fitting model than the NAD

estimation, and replicated quite well. Even after considering additional parameters estimated for multinomial distribution, the An Information Coefficient AIC (Akaike, 1985) supports the superior fit of the MNAD estimation for data sets 2 through 6. Therefore, the MNAD estimates should be used. But let us consider the following observation. First of all, the item response model is truly a one parameter logistic model regardless of the ability distribution, i.e., parameters of correct model is being estimated. Data set one is from the normal distribution, and the NAD estimation is using the correct distribution. Despite of the correct item response model and ability distribution being used for NAD estimation on data set one, the value of -2log-likelihood does not indicate the best model fit. Frequently the wrong item response model and ability distribution fits "better" to the model. It is clear that -2log-likelihood is not a good indicator of the model fit. In light of this, the MNAD does not have a clear advantage over the NAD estimation in terms of model fit.

Study 2:

The second study focused on the effect of varying the ability distribution on the item parameters and the posterior ability distribution. It is clear that when the true ability distribution is different from the prior distribution, the estimated model parameters would be biased. We decided to investigate the degree of skewness of ability distribution and the amount of bias in the IRT item parameter estimation as it applies to the analysis of NAEP data.

In NAEP assessments, over-sampling of minority populations is done to ensure the precision of estimates of performance for the subpopulations. Because of this over-sampling, the weights are necessary to achieve the proper proportion of the subgroup in the population. Typically, the weights are not available until several months after the assessments are administered. In the past, the item analyses and the item parameter estimations were performed using the unweighted data. It was thought that weights affect item parameter estimation very little, and that the effects can be eliminated by a conventional linear transformation. However, the final analyses are made using the post-stratified weights so that the representability of the national population can be maintained. As a result, two different ability distributions are used, one for estimation and another for reporting. This study addresses the issue of whether or not the use of unweighted data for item parameter estimation produces ses in the final estimates of subgroup proficiency and, if so, the magnitude of that bias.

This study considers the more realistic range of ability distributions for the NAEP

assessment. In a random sample of 2000 examinees from the 1986 and 1988 age 13 mathematics trend assessments, 1391 were white and 609 were non-white. All examinees in this sample received an identical booklet consisting of three blocks of items from mathematics, science, and reading. The mathematics block being studied here contained 37 items and presented aurally using a tape recorder. Five combinations of weights were used A:(1.0, 0.0), B:(1.0, 1.0), C:(1.15, 0.657), D:(1.293, 0.328) and E:(0.0, 1.0) for the white and the non-white samples respectively. These weights were uniformly applied to all samples in a given subpopulation.

Five separate BILOG runs were performed to obtain five sets of the two parameter IRT item parameters, each set uniquely corresponding to the one of the five weights combinations. The number of quadrature points were set to 20, and the standard normal prior distribution was used. Based on the estimated item parameters, the posterior ability distributions of the subpopulations were each calculated separately.

Results (study 2)

Five sets of estimated item parameters are presented in Table 2, and exemplary three sets of parameters are plotted in Figures 7 and 8. It is evident that the general direction of estimated values for the weight combinations A to E as the weights shift. The absolute difference of the slope parameter estimates vary upward of .3 across the five weight combinations. The most extreme combinations of A and E differed most; other combinations are found in between the two. The results are quite similar for the threshold parameters as well. The weights A and E differed by as much as .4. The difference on the estimated item parameters between the A and E weights combinations are frequently five times larger than standard errors. Clearly, the standardizing (for A through E, mean = 0.0 and s.d. = 1.0) ability distributions being different influence the item parameters. However, the fact that the deviations are not in one direction indicate that some other factor besides the ability distribution at work. At an item level, the different weight combinations result in significant differences on the estimated item parameters.

Even though the estimated parameters may differ a great deal at an item level depending upon the weights being used, the test as a whole may function equivalently regardless of the weights. In order to answer this question, posterior ability distributions were estimated. Table 3 presents the results.

In order to make the five scales comparable, a following method of standardization was applied. First, five sets of the posterior ability distributions for each subpopulation were calculated based upon the five sets of estimated item parameters. Secondly, the means and standard deviations were set at 0.0 and 1.0 by a linear transformation for each scale separately for the weighted sum of the posterior distributions using the true proportion of the population as the weights. Figure 10 presents the standardized posterior distributions of two subpopulations from the two most extreme weights combinations, A(1.0,0.0) and E(0.0,1.0). It is quite remarkable that the two posterior distributions from different sets of item parameters are very similar to each other once a standardization is applied. Not only the correct mean and standard deviations are obtained but also the oddity of the shapes themselves are captured by seemingly very different item parameters.

Discussion

This study supports the fact that the prior distribution in MML estimation should be as close to reality as possible. Because of the lack of knowledge of the real ability distribution, we may be compelled to estimate the multinomial ability distribution simultaneously with the item parameters. When applied on real data, this method is often unstable and sensitive to the initial values. Out of this study a very practical procedural change has been made. The following two step procedure is now instituted. At first, find converged model parameters using the normal prior distribution. Using those converged values as initial values, continue to estimate the multinomial ability distribution.

The potential contributions of this study are three fold. First, it sheds strong doubt on the presumed nature of the IRT scale being an interval scale. Secondly, it makes a practical recommendation regarding an IRT parameter estimation procedure when the ability distribution is non-normal. Thirdly, this study is applicable to a situation when it is necessary to equate multiple sets of item parameters which are estimated using samples from different ability distribution, such as differential item functioning (DIF) study. Frequently, the model-based DIF study focuses on the item characteristic curve at an item level without considering the possibility of the real consequence on the ability estimates being negligible.

References

Akaike, H (1985) Prediction and entropy. In A. C. Atkinson & S. E. Fienberg (Eds.), *A Celebration of Statistics* (pp. 1-24). New York: Springer-Verlag.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical Theories of Mental Test Scores* (pp. 397-472). Reading MA: Addison-Wesley.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika*, 46, 443-459.

Braun, H. I. (1988) A new approach to avoiding problems of scale in interpreting trends in mental measurement data. *Journal of Educational Measurement*, 25, 171-191.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Erlbaum.

Mislevy, R. J. (1987). Recent developments in IRT in E. Z. Rothkoph, Ed., *Review of Research in Education* (Vol. 14, pp. 239-275). Washington, DC: American Educational Research ASsociation.

Mislevy, R. J., & Bock, R. D. (1982). *BILOG*: Maximum likelihood item analysis and test scoring with logistic models for binary items. Chicago: International Education Services.

Ramsay, J. O. (1989). Maximum marginal likelihood estimation for semiparametric item analysis. Unpublished manuscript, McGill University, Canada.

Ramsay, J. O., & Abrahamowicz, M. (1989) Binomial regression with monotone splines: a psychometric application, Unpublished manuscript, McGill University, Canada.

Table 1

The Model Fit Index of Two Estimation Methods

Data	Replication	Population Distribution	
		Normal	Multinomial
1	1	73040.6	73030.9
2	1	73214.2	73125.7
3	1	73704.7	73456.0
4	1	73124.7	72712.2
5	1	73412.3	72729.4
6	1	73705.4	72781.6
1	2	73452.9	73441.3
2	2	72709.9	72631.4
3	2	73260.7	73001.3
4	2	73096.2	72670.9
5	2	73205.7	72580.8
6	2	72946.7	72668.0

Table 2

Standardized Estimated Item Parameters

Item	White only		10% NW		20% NW		30% NW		Non-White	
	A	B	A	B	A	B	A	B	A	B
1	1.91	-0.85	1.85	-0.93	1.85	-0.96	1.86	-0.99	1.83	-1.12
2	1.38	-1.03	1.50	-0.91	1.56	-0.87	1.58	-0.88	1.66	-0.96
3	0.93	-1.61	0.98	-1.72	1.08	-1.75	1.15	-1.73	1.35	-1.70
4	0.94	-1.62	0.91	-1.79	0.86	-2.02	0.86	-2.06	0.90	-2.14
5	0.61	-0.99	0.70	-0.85	0.84	-0.73	0.89	-0.72	1.04	-0.76
6	0.89	-1.46	0.99	-1.39	1.17	-1.28	1.25	-1.26	1.44	-1.27
7	0.75	-0.83	0.90	-0.67	1.10	-0.59	1.19	-0.58	1.44	-0.63
8	1.01	-0.89	1.16	-0.76	1.37	-0.68	1.45	-0.68	1.68	-0.74
9	1.17	-0.54	1.15	-0.52	1.02	-0.56	0.98	-0.59	0.95	-0.70
10	1.14	-0.33	1.25	-0.27	1.37	-0.25	1.41	-0.27	1.58	-0.37
11	0.63	-1.57	0.64	-1.59	0.68	-1.59	0.71	-1.58	0.79	-1.58
12	1.22	-0.08	1.31	-0.04	1.37	-0.03	1.39	-0.05	1.45	-0.16
13	1.50	-0.66	1.51	-0.66	1.49	-0.67	1.49	-0.70	1.52	-0.80
14	1.30	-0.12	1.24	-0.17	1.19	-0.21	1.18	-0.24	1.19	-0.36
15	0.90	0.04	0.72	-0.13	0.61	-0.31	0.58	-0.38	0.53	-0.57
16	1.01	-0.04	0.98	-0.05	0.93	-0.07	0.90	-0.10	0.85	-0.21
17	0.87	-0.70	0.90	-0.57	0.83	-0.52	0.80	-0.54	0.73	-0.64
18	1.17	0.23	1.31	0.32	1.42	0.39	1.44	0.39	1.47	0.30
19	0.59	-0.58	0.52	-0.82	0.45	-1.10	0.44	-1.17	0.41	-1.46
20	0.16	3.35	0.16	3.41	0.12	4.23	0.10	4.99	0.06	8.74
21	1.79	0.07	1.69	0.02	1.72	-0.03	1.73	-0.07	1.80	-0.21
22	1.41	-0.91	1.34	-0.89	1.24	-0.93	1.21	-0.96	1.15	-1.12
23	1.65	0.30	1.78	0.35	1.80	0.39	1.78	0.38	1.76	0.29
24	0.76	0.60	0.60	0.55	0.46	0.53	0.42	0.53	0.31	0.53
25	0.81	0.06	0.92	0.17	1.00	0.23	1.01	0.22	0.99	0.13
26	0.70	0.67	0.71	0.62	0.80	0.52	0.83	0.46	0.90	0.28
27	0.93	0.42	0.90	0.40	0.92	0.37	0.92	0.34	0.93	0.21
28	1.22	0.94	1.27	0.90	1.39	0.81	1.44	0.76	1.61	0.56
29	1.71	0.42	1.73	0.40	1.77	0.37	1.78	0.34	1.84	0.20
30	1.13	0.41	1.07	0.37	1.04	0.34	1.05	0.31	1.05	0.17
31	1.19	0.69	1.22	0.70	1.27	0.71	1.28	0.69	1.27	0.59
32	0.86	0.96	0.73	0.98	0.59	1.07	0.54	1.12	0.43	1.26
33	0.89	0.47	0.90	0.46	0.92	0.44	0.91	0.42	0.92	0.29
34	1.27	1.06	1.15	1.02	1.14	0.92	1.16	0.86	1.22	0.66
35	0.63	1.55	0.70	1.50	0.77	1.43	0.79	1.39	0.82	1.24
36	1.40	1.26	1.26	1.23	1.24	1.12	1.26	1.06	1.32	0.84
37	1.22	1.31	1.07	1.31	0.94	1.32	0.90	1.30	0.83	1.23

Standardized based on the sum of log a =0.0 and the sum of b = 0.0

Table 3

Posterior Means and Standard Deviations
Estimated by RESOLVE

Parameter Sets(w,nw)[1]	Sub-Population	Proportion of non-white population 0.3 Mean (sd)	0.2 Mean (sd)	0.1 Mean (sd)	Stabdardized Mean (sd)
1(.0,1.)					
	Total		.780 (1.18)		.0 (1.00)
	White		.946 (1.16)		.141 (0.98)
	N-white		.118 (1.01)		-.561 (0.85)
2(.7,.3)					
	Total	.073 (1.01)	.147 (1.01)	.219 (1.01)	.0 (1.00)
	White		.289 (1.00)		.142 (0.99)
	N-white		-.424 (0.85)		-.571 (0.84)
3(.8,.2)					
	Total	.017 (1.01)	.088 (1.01)	.159 (1.00)	.0 (1.00)
	White		.230 (0.99)		.142 (0.98)
	N-white		-.479 (0.85)		-.567 (0.84)
4(.9,.1)					
	Total	-.046 (1.01)	.025 (1.00)	.096 (1.00)	.0 (1.00)
	White		.167 (0.99)		.142 (0.99)
	N-white		-.543 (0.85)		-.568 (0.85)
5(1.,.0)					
	Total		-.062 (0.99)		.0 (1.00)
	White		.082 (1.01)		.144 (1.02)
	N-white		-.638 (0.86)		-.576 (0.87)

[1]Sets of estimated item parameters on data with various proportion of white and non-white subpopulation (w,nw).

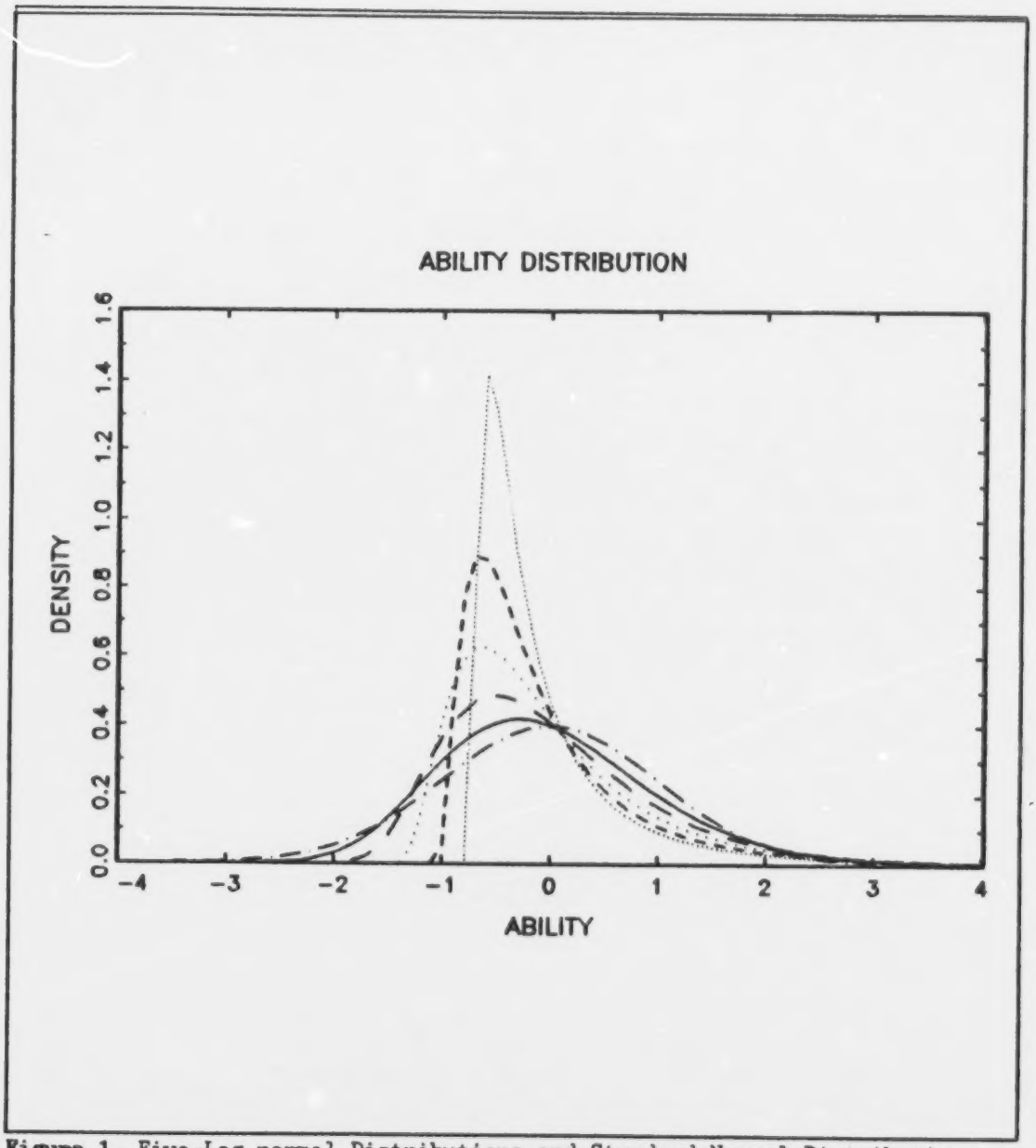

Figure 1 Five-Log normal Distributions and Standard Normal Distribution

1 normal

2
3
4
5
} lognormal

6

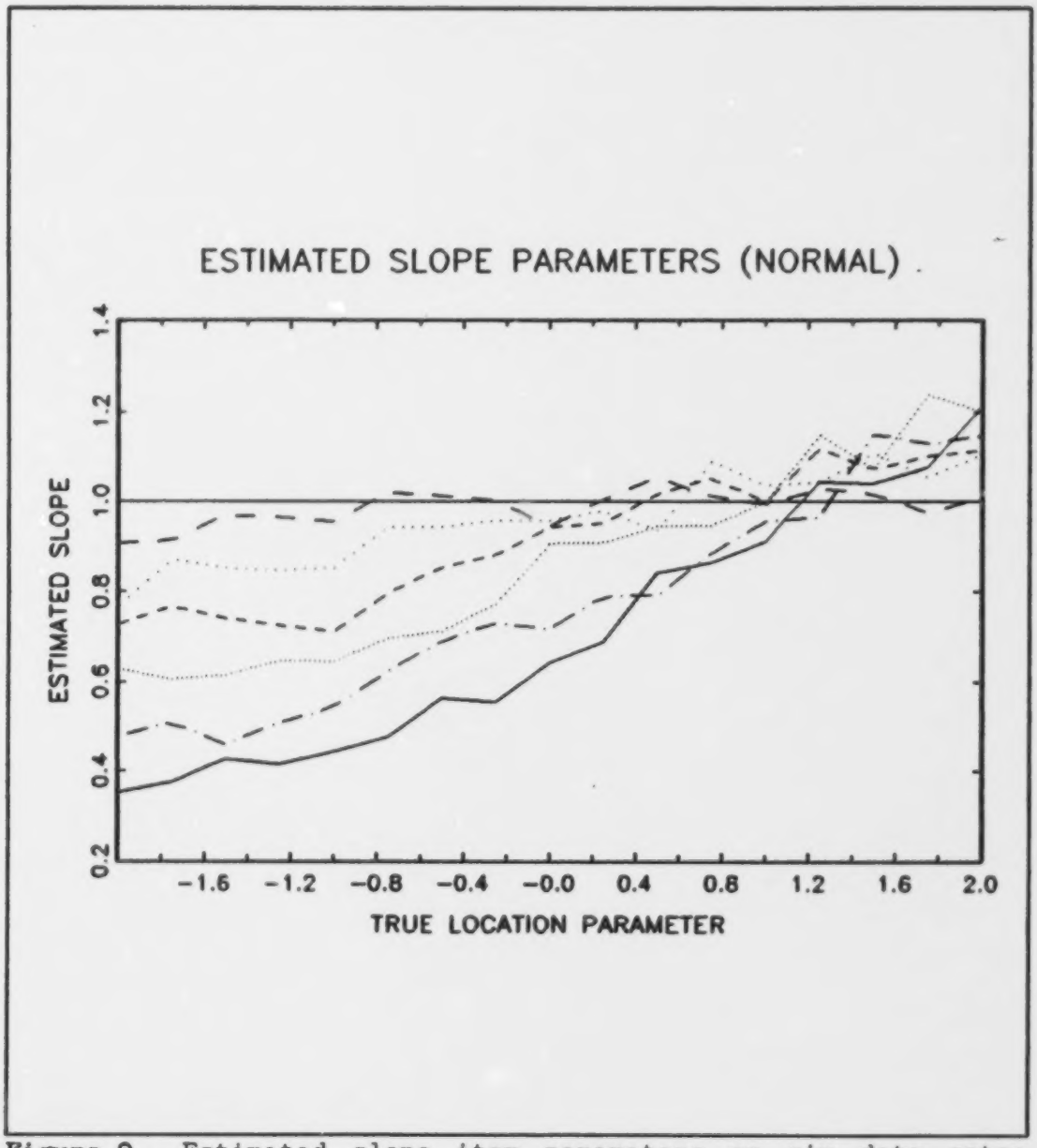

Figure 2 Estimated slope item parameters on six data sets: population distribution fixed as normal

---- 1
..... 2
------ 3
.... 4
- - - 5
——— 6

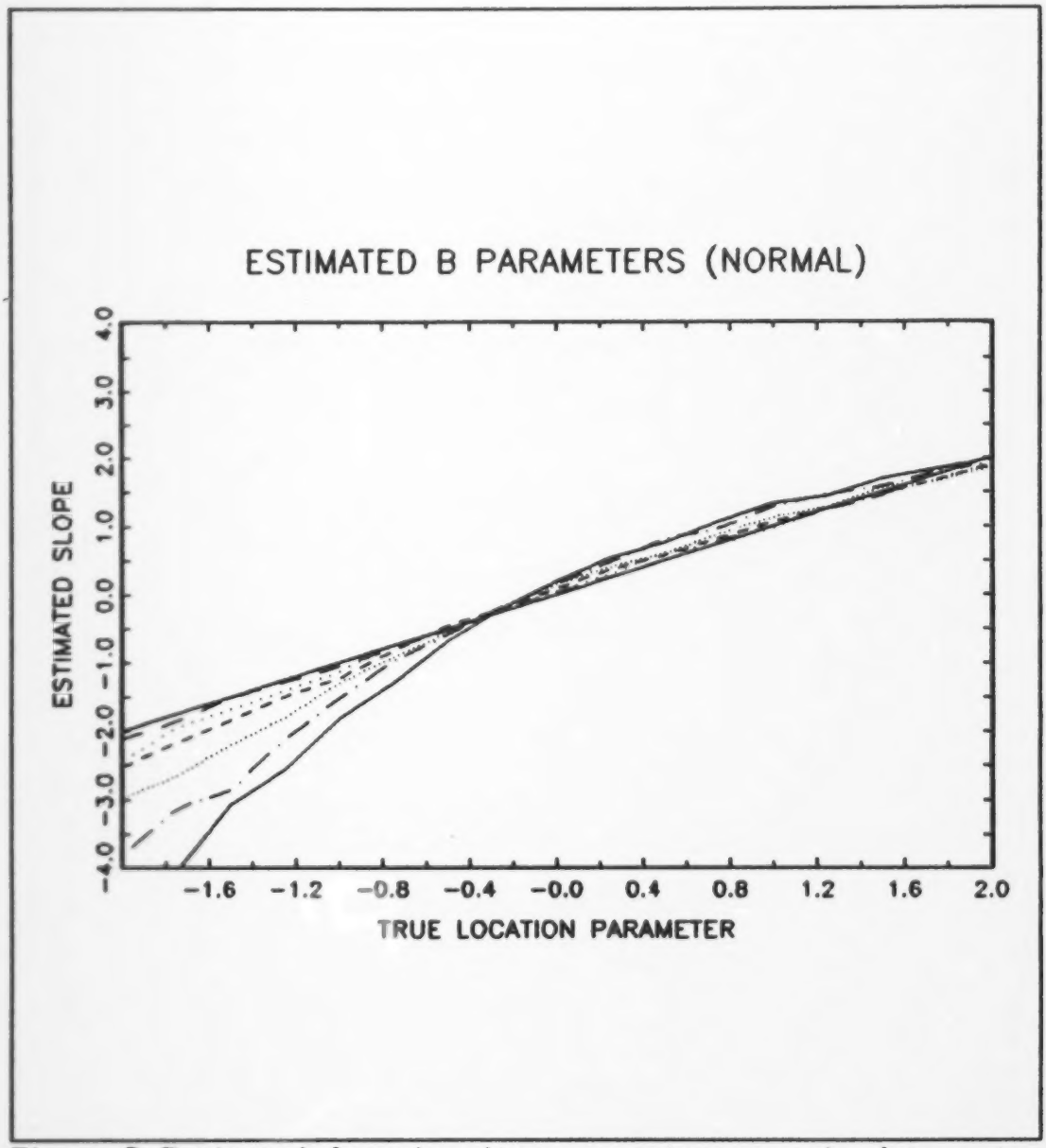

Figure 3 Estimated location item parameters on six data sets: population distribution fixed as normal

---- 1
- - - - - 2
------ 3
..... 4
-·-·-5
——— 6

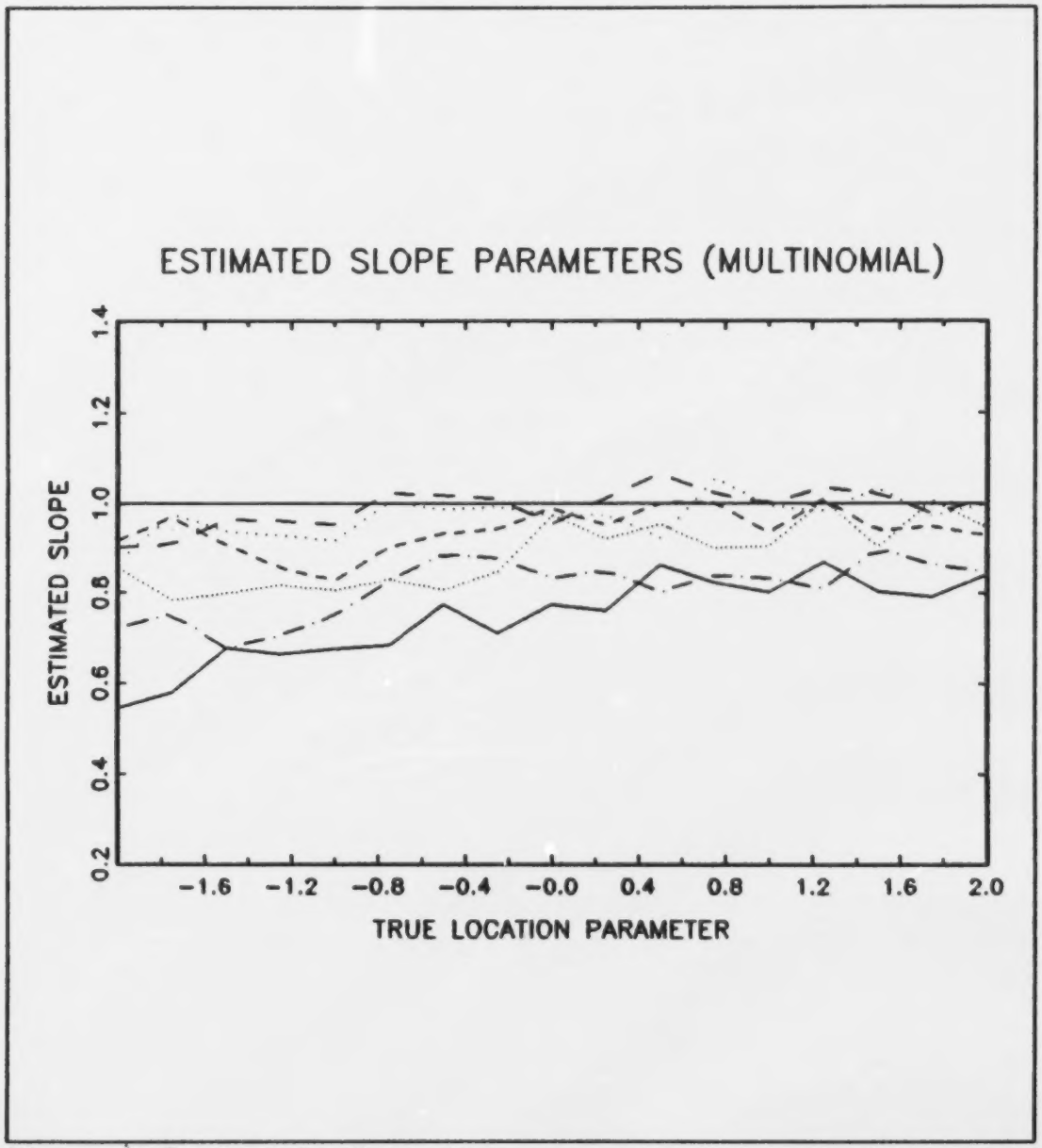

Figure 4 Estimated slope item parameters on six data sets: multinomial population distribution estimated

1
2
3
4
5
6

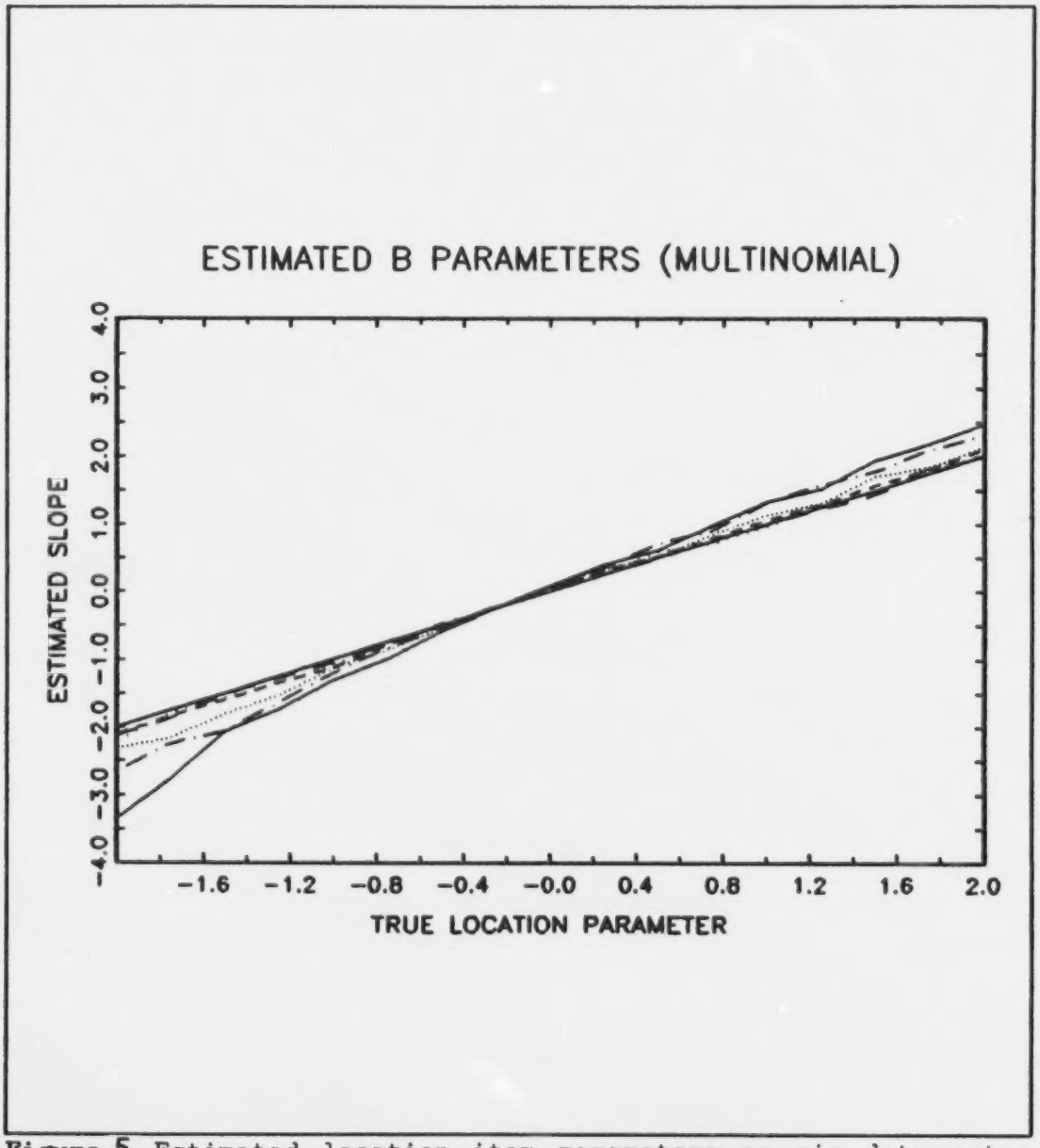

Figure 5 Estimated location item parameters on six data sets: multinomial population distribution estimated

1
2
3
4
5
6

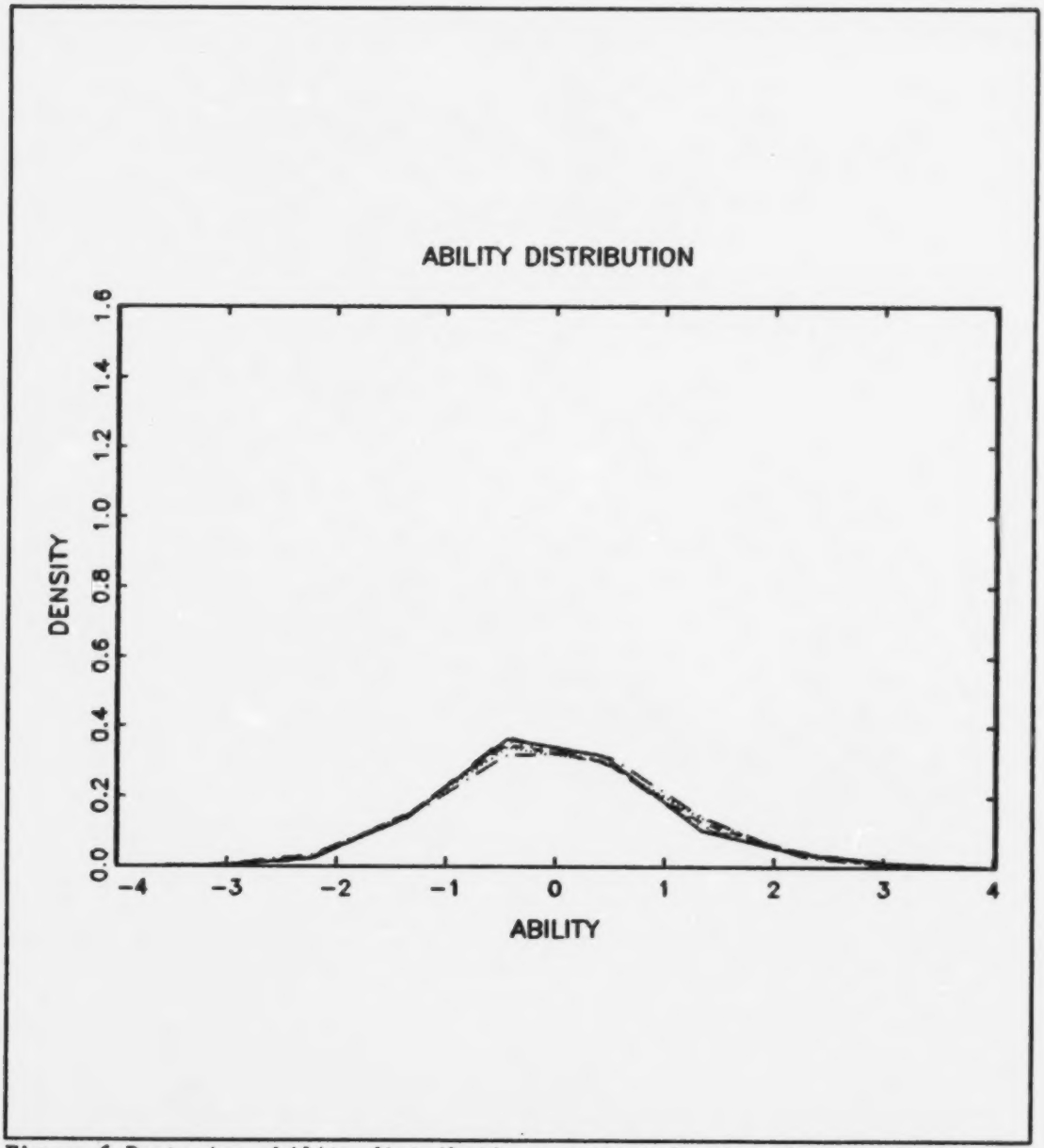

Figure 6 Posterior ability distributions estimated using the normal prior

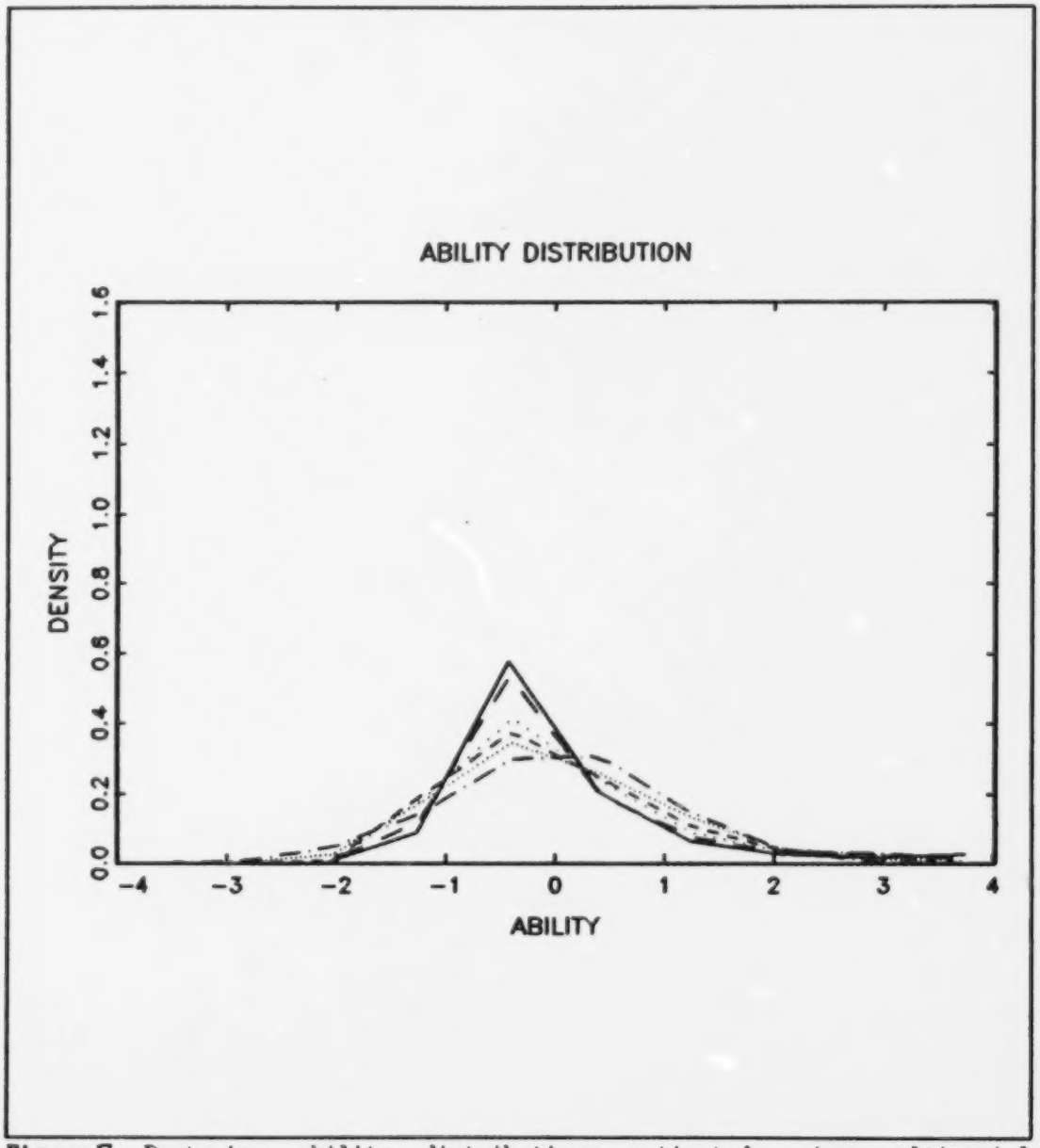

Figure 7 Posterior ability distributions estimated using multinomial distribution

1
2
3
4
5
6

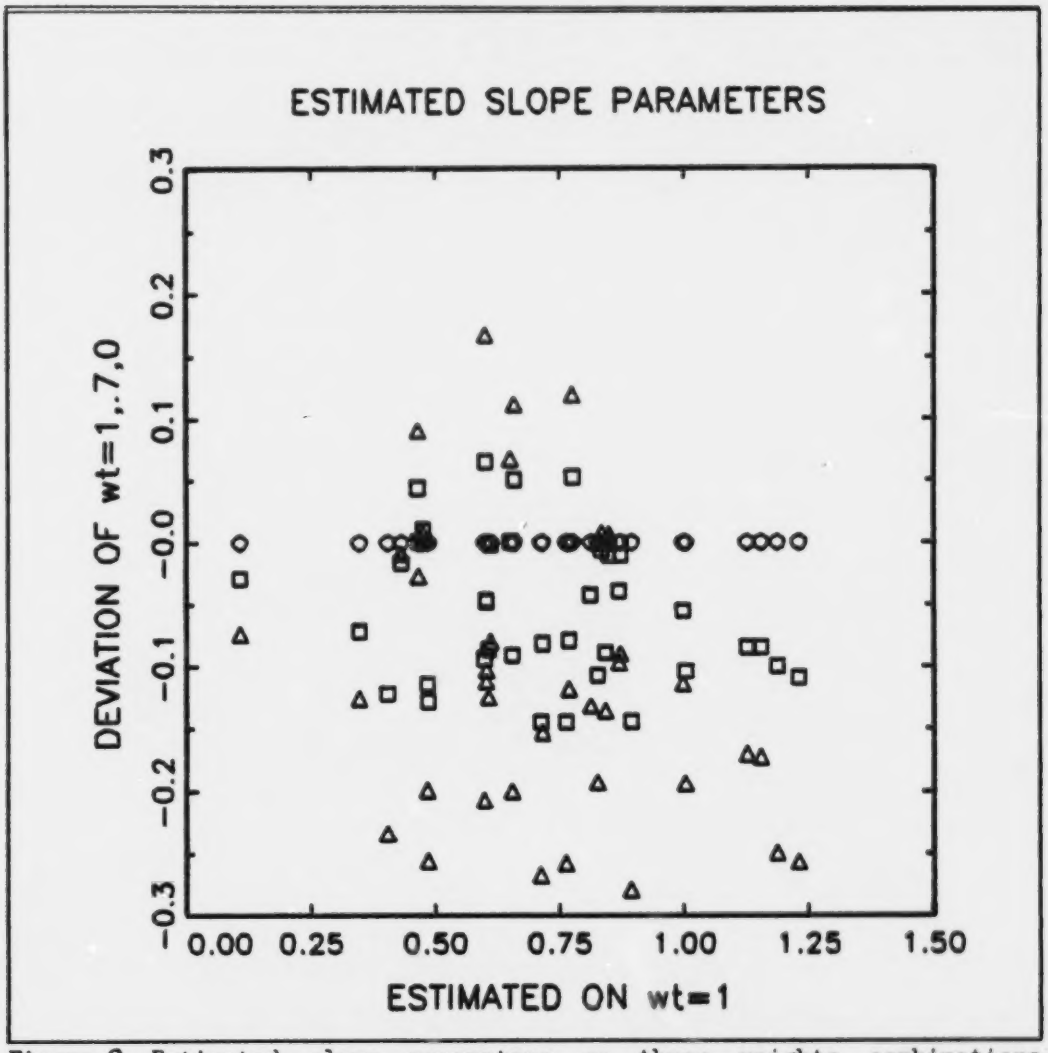

Figure 8 Estimated slope parameters on three weights combinations (1.0,0.0), (0.7,0.3), and (0.0,1.0) for white and non-white populations

○ (1.0, 0.0)
□ (.7, 0.3)
△ (0.0, 1.0)

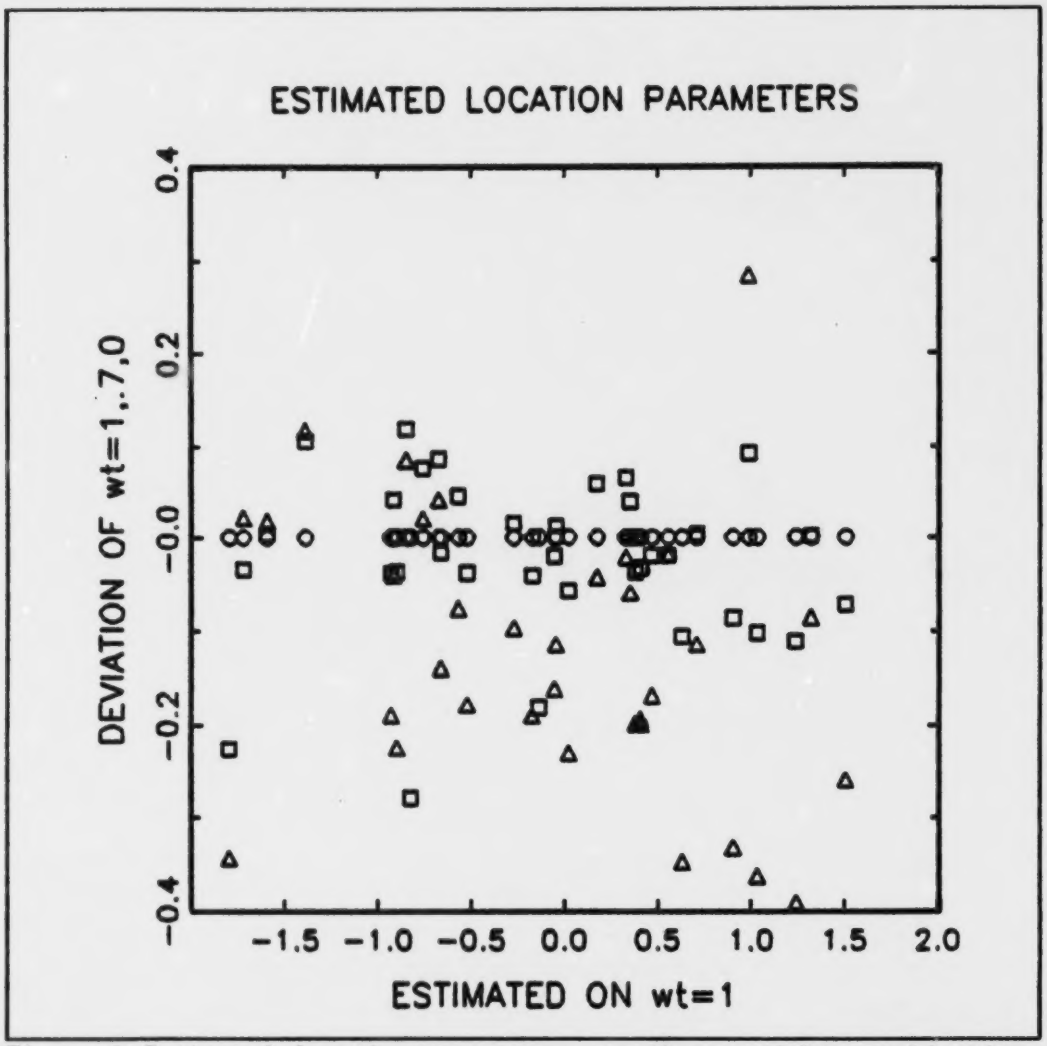

Figure 9 Estimated location parameters on three weights combinations (1.0,0.0), (0.7,0.3), and (0.0,1.0) for white and non-white populations

○ (1.0, 0.0)

□ (.7, .3)

△ (0.0, 1.0)

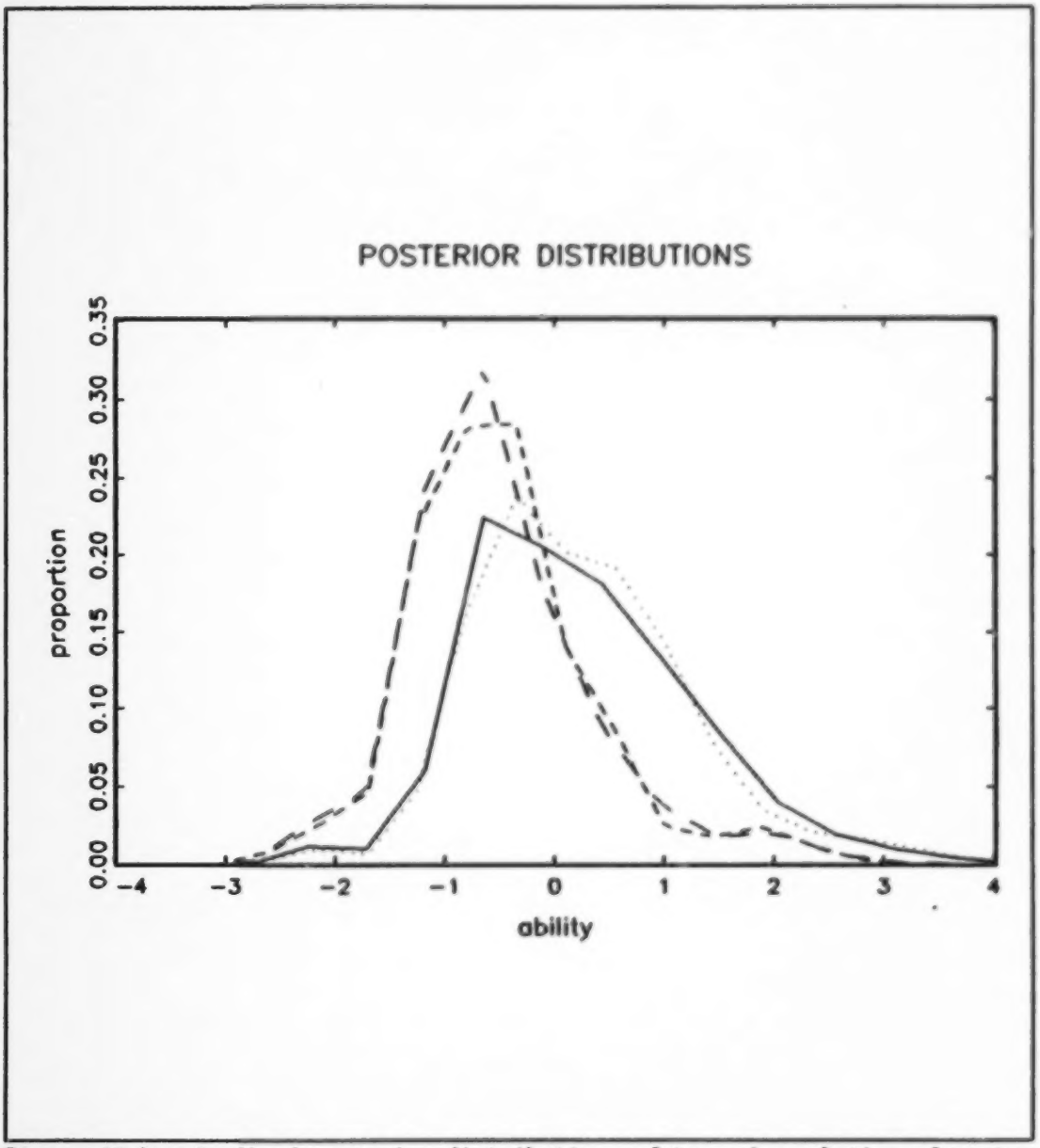

Figure 10 Standardized posterior distributions of two subpopulations from two weights combinations A(1.0,0.0) and E(0.0,1.0) for white and nonwhite subpopulations

——— White

– – – Nonwhite